DOCUMENT RESUME

ED 431 018 TM 029 850

AUTHOR	Parkes, Jay T.; Suen, Hoi K.; Zimmaro, Dawn M.; Zappe, Stephen M.
TITLE	Structural Knowledge as a Pre-requisite to Valid Performance Assessment Scores.
PUB DATE	1999-04-00
NOTE	28p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, Canada, April 20-22, 1999).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*College Students; *Concept Mapping; Higher Education; Knowledge Level; *Performance Based Assessment; Scores; *Transfer of Training; *Validity
IDENTIFIERS	*Structural Knowledge

ABSTRACT

Structural knowledge is a pre-requisite to valid performance assessment scores because structural knowledge leads to better transfer. Better transfer will cause consistency to occur between task performances for a given individual, and this consistency allows for less construct-irrelevant variance and more construct-relevant variance to be present in the scores. This means more valid scores. This paper provides empirical support for each link in this argument. Concept maps are used as representations of the structural knowledge taught in a Political Science I course. Students in that course (approximately 325 students) were assessed using an essay performance assessment. Those students in the concept-mapping condition are shown to write more consistent essays than those who do not use a concept map while writing. That is, the amount of task-related variance is related to the immediacy of having constructed a concept map. Structural knowledge is, therefore, a pre-requisite to valid performance assessment. (Contains 1 figure, 5 tables, and 88 references.) (Author/SLD)

TM

Structural Knowledge as a Pre-requisite to Valid Performance Assessment Scores

Jay T. Parkes

University of New Mexico

Hoi K. Suen

Dawn M. Zimmaro

Stephen M. Zappe

The Pennsylvania State University

Paper presented at the Annual Meeting of the National Council on Measurement in Education: Montreal, Canada. April, 1999. Correspondence regarding this paper should be addressed to: Jay Parkes, Educational Psychology Program, 128 Simpson Hall, University of New Mexico, Albuquerque, NM 87131-1246. E-mail: parkes@unm.edu

Abstract

Structural knowledge is a pre-requisite to valid performance assessment scores because structural knowledge leads to better transfer. Better transfer will cause consistency to occur between task performances for a given individual and this consistency allows for less construct-irrelevant variance and more construct-relevant variance to be present in the scores. This means more valid scores.

This paper provides empirical support for each link in this argument. Concept maps are used as representations of the structural knowledge taught in a Political Science 1 course. Students in that course are assessed using an essay performance assessment. Those students in the concept-mapping condition are shown to write more consistent essays than those who do not use a concept map while writing. That is, we demonstrate that the amount of task-related variance is related to the immediacy of having constructed a concept map. Therefore structural knowledge is a pre-requisite to valid performance assessment.

Structural Knowledge as a Pre-requisite to Valid Performance Assessment Scores

Since performance assessments tend to include only a small sampling of tasks, the weight of content sampling and construct representation is heavily borne by those few tasks. The ability to generalize from the performance on those few tasks to a broader domain hinges on the degree to which the examinee's performances on those two tasks are consistent. If the performances across tasks are highly consistent, generalizable inferences seem appropriate whereas if the performances are not consistent, then such generalizations are not appropriate. This generalizability is an issue of both reliability and validity. It is a reliability issue insofar as consistent performances along with congruent ratings will produce consistent, stable scores. It is a validity issue insofar as methodological concerns can be dismissed and construct issues considered, that is, there is more construct-relevant variance and less construct-irrelevant variance. At the moment, performance assessments routinely seem to deliver neither.

The Problem of Task Related Variance in Performance Assessments

A near consensus of opinion holds that reliable scores are difficult to get from performance assessments. Indeed, low reliability coefficients are seen as one of the major roadblocks, if not the pre-eminent roadblock, to the implementation of large-scale, high stakes performance assessments (Burger & Burger, 1994; Dunbar, Koretz, & Hoover, 1991; Herman, 1991; Linn, 1993, 1994; Linn, Baker, & Dunbar, 1991; Mehrens, 1992; Messick, 1994; Shavelson, Baxter & Pine, 1992). Low coefficients have been shown in numerous and diverse subject matters and populations such as behavioral observation in pre-school (McWilliam & Ware, 1994); middle school science (Shavelson & Baxter, 1992; Shavelson, Baxter & Gao, 1993); secondary school writing and mathematics (Koretz, Klein, McCaffrey, & Stecher, 1993; Koretz, Stecher, Klein & McCaffrey, 1994; Koretz, Stecher, Klein, McCaffrey & Deibert, 1993) and college writing (Nystrand, Cohen, & Dowling, 1993).

The most problematic source of error variance – and hence, low reliability coefficients -- is that of task and subject-by-task variance (Brennan, 1996; Brennan & Johnson, 1995; Green, 1995; Linn, 1993; Linn, 1994; Linn, Baker, & Dunbar, 1991; Swanson, Norman & Linn, 1995). In literally dozens of studies covering many diverse populations and subject areas, the lack of agreement between different tasks on an assessment has been documented (Baker, 1992; Breland, Camp, Jones, Morris, & Rock, 1987; Brennan, Gao & Colton, 1995; Coffman, 1966; College

Board, 1988; Dunbar, Koretz & Hoover, 1991; Gamache & Brennan, 1994; Gao & Colton, 1996; Gao, Shavelson, & Baxter, 1994; Koretz, Stecher, Klein, McCaffrey & Deibert, 1993; Lane, Liu, Ankenman, & Stone, 1996; Lane, Stone, Ankenman & Liu, 1994; Linn, 1993; Mehrens, 1992; Miller & Legg, 1993; Shavelson & Baxter, 1992; Shavelson, Baxter & Gao, 1993; Shavelson, Baxter & Pine, 1990; Shavelson, Baxter & Pine, 1992; Swanson, Norcini & Grosso, 1987).

From the perspective of construct-relevant and construct-irrelevant variance (Messick, 1994, 1995), these task-related variances are also problematic. Variance associated with tasks is considered construct-irrelevant variance, and thereby reduces the validity of the scores from the assessment.

Therefore, if performance assessments are ever going to be of use where high-quality scores are required, the issue of task-related variance needs to be addressed. That is, ways must be found to reduce task and subject-by-task variance components (Linn & Burton, 1994): But in order to do that, some explanation for task variance needs to be offered and some solution based on that explanation attempted. Since all assessments are fundamentally cognitive acts, or more specifically are acts of problem-solving (Snow & Lohman, 1989), perhaps an explanation lies in cognitive psychology. As Snow & Lohman (1989) state: "... new attempts to assess the more dynamic properties of task performance, understood in modern information-processing terms, might well pay off (p. 316)."

Task-Related Variance as a Transfer Problem

Task variance and task-by-person variance are both problematic because, according to the generalizability sampling framework, they should be minimal. But they are quite common as has been previously cited. To put it another way, someone can be asked to do two or more different tasks and get two or more different results. This sounds very similar to research on the problem of transfer of learning. As Linn, Baker & Dunbar (1991) observe: "The limited generalizability from task to task is consistent with research in learning and cognition that emphasizes the situation and context-specific nature of thinking (p. 19)." It is in the transfer literature that a possible explanation for task-related variance can be found.

The issues of transfer and analogical reasoning have vexed and perplexed educators and cognitive psychologists for a long time (cf. Lave, 1988). Students often can do a set of problems and seemingly understand a principle, yet they fail to employ that principle or solution in a novel situation. It seems reasonable to expect students to be able to do that, especially since transferring knowledge is an everyday part of life (Perkins &

Salomon, 1988), and yet much research exists to show that they often fail miserably at it. It sounds very similar to the problem of task-related variance: differential performance on tasks where none is anticipated.

Analogical reasoning, one form of transfer, has long been the focus of much cognitive psychology research. It has been defined as "fundamentally the ability to utilize a well understood problem to provide insight and structure for a less understood problem (Grandgenett & Thompson, 1991; p. 294)." Gick & Holyoak (1983), who performed some of the classic work on analogical reasoning, define it thus: "The essence of analogical thinking is the transfer of knowledge from one situation to another by a process of *mapping* -- finding a set of one-to-one correspondences (often incomplete) between aspects of one body of information and aspects of another (p. 2)." The conditions under which such a match is made are not clearly understood, but different suggestions have been made. Gick & Holyoak (1980) discuss the importance of the level of macrostructure at which this mapping takes place. That is, the mapping might be happening at a broad or general enough level as to be useless or, conversely, at a level too detailed to be useful.

A typical transfer situation may consist of the following steps (Brown & Clement, 1989; Gick & Holyoak, 1980; Novick, 1988; Perkins & Salomon, 1989; Pierce, Duncan, Gholson, Ray, & Kamhi, 1993). First, the base situation must be completely understood and its features and solution learned in generalizable terms. Second, the common elements between the base and the target must be noticed. Third, the relevant and useful aspects of the base must be extracted from memory. Fourth, the same relationships and solutions which worked in the base must be applied successfully to the target. This process is further complicated, however, as the base and target become more and more dissimilar. In that case, several attributes in the base are irrelevant to the target and may actually be detrimental to solving the target situation (Novick, 1988). In this case, learners are presented with the additional challenge of deciding which aspects of the base to use and which to ignore.

Transfer is further complicated because the base and target, as well as the activity of transfer itself, are situated in some context which could very well alter the difficulty of making the transfer (Bassock & Holyoak, 1989; Gick & Holyoak, 1980; Lave, 1988). Perkins & Salomon (1989) note that transfer is "highly specific and must be cued, primed, and guided (p. 19)." Garner (1990) comments that little attention has been paid to this issue of context when addressing the issue of strategy use.

In essence, analogical reasoning is an indefinable and subtle balance between general principles or knowledge and specific situations. It is a balance because, if all of a learner's knowledge is situationally specific, no

abstraction will take place and transfer will not occur (e.g. Bassock, 1990; Bassock & Holyoak, 1989; Garner, 1990; Perkins & Salomon, 1989); conversely, if knowledge is very general, too few features exist with which to make matches, and again, transfer does not occur (Perkins & Salomon, 1989). Thus effective transfer or analogical reasoning is composed of both general and context-specific elements (Phye, 1989; Perkins & Salomon, 1989). The final complication to the process is that the learner must make the connection between the two situations somehow.

Given these conditions, it seems surprising that transfer ever occurs. Indeed, it fails a great deal of the time and for various reasons. One possibility is that students fail to use consistent approaches to problems, perhaps due to inadequate learning of the base situation (Perkins & Salomon, 1988; Siegler, 1989). Ruiz-Primo, Baxter, & Shavelson (1993) report that students used different procedures to attack the same problems on different occasions. A second reason for failure to transfer depends on the type of problems being solved. Ill-defined problems -- those with fuzzy parameters -- provide more of a challenge to transfer than do well-defined problems because the mapping or matching process between the target and base is confounded. These types of problems are more common in the social sciences (Voss, 1988). A third type of transfer failure occurs when students are instructed only on general principles and see no concrete or context-specific examples of the principles in action (Bransford, Franks, Vye, & Sherwood, 1989; Bransford, Sherwood, Vye, & Rieser, 1986; De Leeuw, 1983; Larkin, 1989; Perkins & Solomon, 1989; Stratton & Brown, 1972).

These descriptions of transfer problems sound exactly like the problems encountered on assessments with large task and task-by-person variance. Green (1995) comments that on such assessments, the tasks have little common variance and much specific variance. This sounds much like Phye's (1989) description of transfer consisting of both general and context-specific elements. To use the analogical reasoning vocabulary, the mapping of attributes between task variance situations and transfer failures seems to indicate that solutions to the base (transfer failures) might well be applicable to the target (task variance in performance assessments). Many different solutions to transfer problems have been attempted which focus on three basic parts of the transfer process. The first class focuses on students' prior knowledge (Bassock & Holyoak, 1989; Brown, Kane, & Echols, 1986; Gentner, 1983; Gentner, 1989; Holyoak, 1984; Mayer & Bromage, 1980; Novick, 1988; Salomon & Perkins, 1989; Voss, 1988; Wideman & Owston, 1991); the second class focuses on inducing a schema, that is, forming a generalizable knowledge base (Cooper & Sweller, 1987; Gick & Holyoak, 1983; Greeno, Moore, & Smith, 1993; Jelsma, Van Merriënboer, & Bijlstra, 1990; Lambiotte & Dansereau, 1992; Novick, 1988; Phye, 1989; Royer, 1979; Royer,

1986; Rumelhart & Norman, 1980; Sweller, 1989; Van Merriënboer & Paas, 1990); and the third focuses on getting students to notice the correspondences between the base and the target (Catrambone & Holyoak, 1989; Gick & Holyoak, 1980, 1983; Phye, 1989; Ross, 1984; Ross, 1987; Salomon & Perkins, 1989; Voss, 1988).

If the problem of task variance is actually a transfer problem, then these solutions should be applicable to it. Their implementation in performance assessment tasks should reduce task variance. Indeed, discussions of the task variance problem have included proposed solutions that sound very much like those proposed for transfer problems. Linn & Burton (1994) acknowledge that prior knowledge deficits might contribute to task variance, and if students have an opportunity to learn the content knowledge as well as gain practice with the task formats, task variance might be reduced. Suen, Sonak, Zimmaro, & Roberts (1997) have proposed a schema induction solution to task variance by hypothesizing that concept maps can be used to bridge the gap between tasks. As Linn & Burton (1994) observe, however, there is some theoretical work but little empirical work here.

Bridging is a technique suggested to promote and teach transfer (Perkins & Salomon, 1988). The technique involves providing some support to help students see the possibility of transfer. Concept mapping is a powerful cognitive tool (Lambiotte & Dansereau, 1992; Novak & Gowin, 1984) and should provide an effective bridge across tasks for a number of reasons. First, research has shown that students with better schema are better able to transfer effectively (Catrambone & Holyoak, 1989; Cooper & Sweller, 1987; Gentner, 1989; Gick & Holyoak, 1983). More broadly, visual representations also aid transfer (Beveridge & Parkins, 1987; Gick & Holyoak, 1980; Gick, 1985; Yang & Wedman, 1993). The use of scaffolded techniques also aids transfer (Choi & Hannefin, 1995; Greenfield, 1984; Harley, 1993; Paas, 1992). Finally, techniques which help students focus on the conceptual structure of knowledge are better at enhancing transfer than techniques that focus on surface features (Bransford, Sherwood, Vye, & Rieser, 1986; Gentner, 1989; Paas, 1992; Weaver & Kintsch, 1992). The present study is designed to benefit from all of these findings.

This investigation brought all of this together empirically to address the issue of task-related variance. The review of literature suggests that task-related variance can be seen as a transfer problem and that solutions to transfer problems which draw on the structural knowledge technique of concept mapping might also address the task variance problem. The main hypothesis for this study is that the task-related variance components for those who constructed and used a concept map while writing will be smaller than the task-related variance components for those who did not construct or use a concept map while writing.

Methodology

This study took place in an introductory Political Science course in American National Government at a large, northeastern university. There were approximately 325 students enrolled in the course. The professor of the course proposes that politics can be thought of as a game with participants, rules, outcomes, resources, etc. It is from this underlying framework (the Play of Power) that the entire course is taught (Eisenstein, Kessler, Williams, & Switzer, 1996). That is, every aspect of American National Government covered in the course is explained using the metaphor. This provides the tight conceptual framework necessary to study issues of transfer.

The primary analytical tool in the course was essay writing. Specifically, students were asked to take a newspaper article and understand and explain the situation in terms of the metaphor. Two of these essays were completed as part of an In-Class Graded Exercise (ICGE). In this study three of these ICGE's are of interest.

Four different treatment conditions were employed in this study. In the concept mapping condition, students were trained in the use of concept mapping as a way to organize the Play of Power metaphor. The "treatment" was that students were asked to create a concept map as a homework assignment and were allowed access to that concept map when working on the ICGE. During the ICGE, students who completed the map and brought it with them were told to refer to the map when formulating their responses.

In the outlining condition, students were also trained to use outlining as a way to organize the course content. As with the first condition, students were asked to produce an outline in advance to which they would be given access during the ICGE. The outlining condition here is a "placebo" control condition. That is, in order to rule out the competing hypothesis that concept mapping reduced variance simply because students had done more preparation for the ICGE's, this condition was added. The intention was that students would spend as much time creating the outlines, but the outlines would not provide the same benefits of enhancing schema formation, providing a visual representation, and improving conceptual understanding that should accrue through the use of concept maps.

The third condition is the "no intervention," or control, condition. Here, neither concept maps nor outlines were required or available to the students during the ICGE.

The fourth condition exists after ICGE 1 and is the previous concept map condition (PCM). These are students who had constructed and used a concept map on a previous ICGE but were not using one on the present ICGE.

The twelve recitation sections of the course were broken into six experimental groups. The treatment conditions were then distributed across the four ICGE's of interest to this study following a multiple baseline design so that, at each ICGE, some students were using a concept map and some were working unaided either as controls or PCM.

The scoring of the ICGE's was completed in three separate scoring sessions, one for each ICGE. The essays were scored for the ability of students to apply the concepts from the Play of Power to the article at hand. Specifically, raters judged the degree to which the student connects, integrates, elaborates and contextualizes the appropriate concepts.

Not all students in each condition completed the homework assignment given to them, so the treatment groups were defined based on what was completed. Also, due to the cost of rating and the commitment of the raters, not every ICGE for every student was scored. Table 1 shows both which cells of the design were scored and what the effective cell sample size was. Since the outlining condition was primarily a form of control group whose functioning is reported elsewhere (Parkes, 1998), those scores will not be of interest in this particular study. Only the Concept Map (CM), Control and PCM conditions are of interest for the first three ICGE's.

Insert Table 1 Here

A separate generalizability study was conducted in each of the nine highlighted cells in Table 1 following this model: $\hat{\sigma}_X^2 = \hat{\sigma}_s^2 + \hat{\sigma}_t^2 + \hat{\sigma}_r^2 + \hat{\sigma}_{st}^2 + \hat{\sigma}_{sr}^2 + \hat{\sigma}_{rt}^2 + \hat{\sigma}_{srt}^2$. The task main effect variance components and the subject-by-task variance components were all compared across different cells. A one-standard error confidence interval was constructed around each estimated component, and these confidence intervals were used to determine if the two components were of statistically significant different sizes. Also, the generalizability coefficients for each cell were calculated and compared.

Results

Table 2 reports the estimated task main effect variance components and their associated standard errors. Table 3 reports the estimated subject-by-task interaction effect variance components and their associated standard errors.

Insert Table 2 Here

Insert Table 3 Here

The main hypothesis was tested through several different comparisons. Table 4 summarizes the comparisons made for the task main effect variance components. No statistically significant differences were observed for the estimated task main effect variance components.

Insert Table 4 Here

The comparisons for the estimated subject-by-task interaction variance components did produce some notable findings, as summarized in Table 5.

Insert Table 5 Here

In two of the six comparisons of changes in the subject-by-task variance components, statistically significant differences were found. The concept mappers' subject-by-task variance component was smaller than the control

group's. In the other four comparisons, statistical significance was not reached, though the relationships among the treatment groups trends according to the predictions made.

From a practical standpoint, however, the real proof is in the effects these different variance components have on the generalizability coefficient. With this particular variance model, the g-coefficient would be given by:

$$\hat{\rho}^2 = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + \frac{\hat{\sigma}_{sr}^2}{n_r} + \frac{\hat{\sigma}_{st}^2}{n_t} + \frac{\hat{\sigma}_{srt}^2}{n_r n_t}}$$

where n_r is the number of raters reading each essay (here, $n_r = 2$); and n_t is the number of tasks, or essays, written by each student (here, $n_t = 2$).

In Figure 1 below, the g-coefficients for both the concept mapping condition and the control condition are given for each ICGE. Note that, for ICGE 3, the comparison is with previous concept mapping condition not the control condition.

Insert Figure 1 Here

In all cases, the g-coefficient is larger for the concept mapping group than for the control or previous concept mapping group. Thus the redistribution of variance due to concept mapping is having a positive effect. It is also noteworthy that the g-coefficients for the concept mapping condition continue to get larger throughout time whereas the control condition g-coefficients remain relatively flat across time.

Conclusions

These results suggest that students who have a concept map of the underlying conceptual framework at their disposal while they write produce more consistent applications of that framework than those students who do not. It is important to note that two effects did not occur. First, there was no effect on the task main effect variance components, only on the subject-by-task variance components. Second, there was no effect of having constructed a map previously but not accessing it while writing.

In retrospect, the lack of a main effect but the presence of an interaction effect is understandable. The effect of the concept map is an individual difference, not a blanket group effect. That is to say that the quality and correctness of the maps and the ability of the student to employ the map while writing would both be different from student to student. That may be why no group effect was seen on the task main effect. As a group, the task variance was not different, but a smaller subject-by-task effect means that within each student, the task variance was less. So each students' own two performances were more consistent. Therefore, it seems quite reasonable to have seen these effects. Having made those qualifications, some drop in task main effect theoretically should have been observed and it was not.

The mixed results regarding the variance components from the Previous Concept Map condition highlight the issue of exactly what the concept maps are doing. There are two potential effects. First, the theoretical argument behind using concept maps was that the maps would strengthen and solidify the students' internal cognitive network of political science concepts. The second effect is that having the concept map to refer to aided the transfer between the framework and the tasks but did not have a strong effect on the students' internal cognitive structure. Since previous concept mappers did not perform as theory predicted when compared to control students, perhaps the second effect is the one observed here. Feedback from the teaching assistants with the course and other research being conducted on the maps themselves seem to indicate that the students did not fully implement concept maps. More specifically, some of the maps were quite poor; and some students expressed confusion about how the maps should help them. It seems more appropriate to make the argument that the maps were helpful "on the spot" but did not have lasting effect as they were implemented. This is not to say that the map was a "cheat sheet." That conclusion is highly unlikely because the map did not contain "answers" or specific facts that would help with application of concepts.

It is important to differentiate between the effects of consistent performances versus the effects of standardized tasks. The maps brought consistency to the students' performances, not to the tasks themselves. The maps seem to have amplified the focus on the application of the underlying framework and de-emphasized the focus on the particulars of a given article as the students wrote. This, then, means that the shift in size of variance components and subsequent shifts in generalizability coefficients is more an issue of validity than reliability. That is to say that the maps did not standardize the task, otherwise, the main effect for task should also have been reduced. Rather the maps reduced the amount of construct-irrelevant variance and increased the amount of construct-relevant

variance in the generalizability model. This is shown in the generalizability coefficients shown in Figure 1. Therefore, there is stronger evidence of construct validity when students employ the concept maps while writing essays.

Although these results are not strong, there is the suggestion that the line of logic used in designing the study seems to still be worth pursuing. There is ample literature and now some empirical findings to support each link in that chain. Structural knowledge, as represented by a concept map, leads to better transfer across tasks. That is, students with a concept map of the underlying framework do less task-specific thinking and more task-general or construct-specific thinking. If students are doing that, then their performances on multiple tasks should be more consistent. In other words, the subject-by-task variance components should be smaller. This in turn increases the proportion of construct-relevant variance and decreases the proportion of construct-irrelevant variance in the total score. And that means more valid scores.

References

Baker, E. L. (1992). *The role of domain specifications in improving the technical quality of performance assessment.* (Tech. Rep.). Los Angeles: UCLA Center for Research on Evaluation Standards and Student Testing. [ERIC Document Reproduction Service No. ED 346 133].

Bassock, M. (1990). Transfer of domain-specific problem-solving procedures. *Journal of Experimental Psychology: Learning, Memory & Cognition, 16 (3)*, 522-533.

Bassock, M. & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15 (1)*, 153-166.

Beveridge, M., & Parkins, E. (1987). Visual representation in analogical problem solving. *Memory & Cognition, 15 (3)*, 230-237.

Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou and A. Ortney (Eds.), *Similarity and analogical reasoning.* (pp. 470-495). Cambridge, NY: Cambridge University Press.

Bransford, J. D., Sherwood, R. D., Vye, N. J., & Rieser, J. (1986). Teaching thinking and problem solving: Research foundations. *American Psychologist, 41 (10)*, 1078-1089.

Bransford, J. D., Sherwood, R. D., Vye, N. J., & Rieser, J. (1986). Teaching thinking and problem solving: Research foundations. *American Psychologist, 41 (10)*, 1078-1089.

Breland, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. A. (1987). *Assessing writing skill* (Research Monograph No. 11). New York: CEEB.

Brennan, R. L. (1996). Generalizability of performance assessments. In G. W. Philips (Ed.), *Technical Issues in Large-Scale Performance Assessment* (pp. 19 - 58). Washington, DC: US Department of Education, Office of Educational Research and Improvement.

Brennan, R. L., & Johnson, E. G. (1995). Generalizability of performance assessments. *Educational Measurement: Issues and Practice, 14 (4)*, 9-12, 27.

Brennan, R. L., Gao, X., & Colton, D. A. (1995). Generalizability analyses of Work Keys Listening and Writing Tests. *Educational and Psychological Measurement, 55 (2)*, 157-176.

Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development, 1*, 103-121.

Brown, D. E., & Clement, J. (1989). *Overcoming misconceptions via analogical reasoning: Factors influencing understanding in a teaching experiment.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. [ERIC Document Reproduction Service No. ED 307 118.]

Burger, S. E. & Burger, D. L. (1994). Determining the validity of performance-based assessment. *Educational Measurement: Issues and Practice, 13 (1)*, 9-14.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15 (6)*, 1147-1156.

Choi, J., & Hannafin, M. (1995). Situated cognition and learning environments: Roles, structures, and implications for design. *Educational Technology, Research and Development, 43 (2)*, 53-69.

Coffman, W. E. (1966). On the validity of essay tests of achievement. *Journal of Educational Measurement, 3,* 151-156.

College Board (1988). *Technical manual for the Advanced Placement program.* New York: Author.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem solving transfer. *Journal of Educational Psychology, 79 (4)*, 347-362.

De Leeuw, L. (1983). Teaching problem solving: An ATI study of the effects of teaching algorithmic and heuristic solution methods. *Instructional Science, 12*, 1-48.

Dunbar, S. B., Koretz, D., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education, 4,* 289-304.

Eisenstein, J., Kessler, M., Williams, B. A., & Switzer, J. V. (1996). *The Play of Power: An Introduction to American Government..* New York: St. Martin's Press.

Gamache, L. M., & Brennan, R. L. (1994, April). *Issues of generalizability: Tasks, raters, and contexts for the NCBE-PT.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Gao, X., & Colton, D. A. (1996, April). *Evaluating measurement precision of performance assessment with multiple forms, raters, and tasks.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Gao, X., Shavelson, R. J., & Baxter, G. P. (1994). Generalizability of large-scale performance assessments in science: Promises and problems. *Applied Measurement in Education, 7 (4),* 323-242.

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Research, 60 (4)*, 517-529.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 77*, 155-170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortney (Eds.), *Similarity and analogical reasoning* (pp. 199-241). Cambridge, England: Cambridge University Press.

Gick, M. L. (1985). The effect of a diagram retrieval cue on spontaneous analogical transfer. *Canadian Journal of Psychology, 39 (3),* 460-466.

Gick, M. L. & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive psychology, 12 (3)*, 306-355.

Gick, M. L. & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Grandgenett, N. & Thompson, A. (1991). Effects of guided programming instruction on the transfer of analogical reasoning. *Journal of Educational Computing Research, 7 (3)*, 293-308.

Green, B. F. (1995). Comparability of scores from performance assessments. *Educational Measurement: Issues and Practice, 14 (4)*, 13-15, 24.

Green, B. F. (1995). Comparability of scores from performance assessments. *Educational Measurement: Issues and Practice, 14 (4)*, 13-15, 24.

Greenfield, P. M. (1984). A theory of the teacher in the learning activities of everyday life. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context.* (pp. 117-138). Cambridge, MA: Harvard University Press.

Greeno, J., Moore, J. & Smith, D. (1993). Transfer of situated learning. In B. Rogoff & J. Lave (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99-167). Norwood, NJ: Ablex.

Harley, S. (1993). Situated learning and classroom instruction. *Educational Technology, 33 (3)*, 46-51.

Herman, J. (1991). Research in cognition and learning: Implications for achievement testing practice. In M. C. Wittrock & E. L. Baker (Eds.). *Testing and cognition* (154-165). Englewood Cliffs, NJ: Prentice Hall.

Holyoak, K. J. (1984). Analogical thinking and human intelligence. In R. J. Sternberg (Ed.) *Advances in the psychology of human intelligence* (Vol. 2, pp. 199-230). Hillsdale, NJ: Erlbaum.

Jelsma, O., Van Merriënboer, J. J. G., & Bijlstra, J. P. (1990). The ADAPT design model: Towards instructional control of transfer. *Instructional Science, 19*, 89-120.

Koretz, D., Klein, S., McCaffrey, D., and Stecher, B. (1993). *Interim report: The reliability of the Vermont portfolio scores in the 1992-93 school year.* (RAND/ RP - 260). Santa Monica, CA: RAND.

Koretz, D., Stecher, B., Klein, S., and McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice, 13 (3),* 5-16.

Koretz, D., Stecher, B., Klein, S., McCaffrey, D., and Deibert, E. (1993). *Can portfolios assess student performance and influence instruction?*. (RAND/ RP-259). Santa Monica, CA: RAND.

Lambiotte, J. G. & Dansereau, D. F. (1992). Effects of knowledge maps and prior knowledge on recall of science lecture content. *Journal of Experimental Education, 60 (3)*, 189 - 201.

Lane, S., Liu, M., Ankenmann, R. D., & Stone, C. A. (1996). Generalizability and validity of a mathematics performance assessment. *Journal of Educational Measurement, 33 (1)*, 71-92.

Lane, S., Stone, C. A., Ankenmann, R. D., & Liu, M. (1994). Reliability and validity of a mathematics performance assessment. *International Journal of Educational Research, 21*, 247-262.

Larkin, J. H. (1989). What kind of knowledge transfers? In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 283-305). Hillsdale, NJ: Erlbaum.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life.* New York: Cambridge University Press.

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis, 15 (1)*, 1-16.

Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher, 23 (9)*, 4-14.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20 (8)*, 15-21.

Linn, R. L. & Burton, E. (1994). Performance-based assessment: Implications of task specificity. *Educational Measurement: Issues and Practice, 13 (1)*, 5-8, 15.

Mayer, R. E., & Bromage, B. (1980). Different recall protocols for technical texts due to advance organizers. *Journal of Educational Psychology, 72*, 209-225.

McWilliam, R. A., & Ware, W. B. (1994). The reliability of observations of young children's engagement: An application of generalizability theory. *Journal of Early Intervention, 18 (1)*, 34-47.

Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice, 11 (1)*, 3-9, 20.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23 (2)*, 13-23.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50 (9)*, 741-749.

Miller, M. D. & Legg, S. M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice, 12 (2)*, 9-15.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* New York: Cambridge University Press.

Novick, L. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14 (3)*, 510-520.

Nystrand, M., Cohen, A. S., & Dowling, N. M. (1993). Addressing reliability problems in the portfolio assessment of college writing. *Educational Assessment, 1 (1)*, 53-70.

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84 (4)*, 429-434.

Perkins, D. N. & Salomon, G. (1988). Teaching for transfer. *Educational Leadership, 46 (1)*, 22-32.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18*, 16 - 25.

Phye, G. D. (1989). Schemata training and transfer of an intellectual skill. *Journal of Educational Psychology, 81 (3)*, 347-352.

Pierce, K. A., Duncan, M. K., Gholson, B., Ray, G. E., & Kamhi, A. G. (1993). Cognitive load, schema acquisition, and procedural adaptation in nonisomorphic analogical transfer. *Journal of Educational Psychology, 85 (1)*, 66-74.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16,* 371-416.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629-639.

Royer, J. M. (1979). Theories of the transfer of learning. *Educational Psychologist, 14,* 53-69.

Royer, J. M. (1986). Designing instruction to produce understanding: An approach based on cognitive theory. In G. D. Phye & T. Andre (Eds.), *Cognitive classroom learning: Understanding, thinking and problem solving* (pp. 83-113). Orlando, FL: Academic Press.

Ruiz-Primo, M. A., Baxter, G. P., & Shavelson, R. J. (1993). On the stability of performance assessments. *Journal of Educational Measurement, 30 (1),* 41-53.

Rumelhart, D. E., & Norman, D. A. (1980). Analogical processes in learning. In J. R. Anderson (ed.), *Cognitive skills and their acquisition.* (pp. 335-359). Hillsdale, NJ: Erlbaum.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24 (2),* 113-142.

Shavelson, R. J., & Baxter, G. P. (1992). What we've learned about assessing hands-on science. *Educational Leadership, 49 (8),* 20-25.

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement, 30 (3),* 215-232.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher, 21 (4),* 22-27.

Siegler, R. S. (1989). Strategy diversity and cognitive assessment. *Educational Researcher,18 (9)* , 15-20.

Snow, R. E. & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 263-331). New York: Macmillan.

Stratton, R. P. & Brown, R. (1972). Improving creative thinking by training in the production and/ or judgment of solutions. *Journal of Educational Psychology, 63,* 390-397.

Suen, H. K., Sonak, B., Zimmaro, D., & Roberts, D. M. (1997). Concept map as scaffolding for authentic assessment. *Psychological Report, 81,* 734.

Swanson, D. B., Norman, G. R., & Linn, R. L. (1995). Performance-based assessment: Lessons from the health professions. *Educational Researcher, 24 (5)*, 5 - 11, 35.

Swanson, D., Norcini, J., & Grosso, L. (1987). Assessment of clinical competence: Written and computer-based simulations. *Assessment and Evaluation in Higher Education, 12*, 220-246.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problemsolving in mathematics and science. *Journal of Educational Psychology, 81*, 457-466.

Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1990). Automation and schema acquisition in learning elementary computer programming: Implications for the design of practice. *Computers in Human Behavior, 6*, 273-289.

Voss, J. F. (1988). Learning and transfer in subject matter learning: A problem solving perspective. *International Journal of Educational Research, 11*, 607-622.

Weaver, C. A., & Kintsch, W. (1992). Enhancing students' comprehension of the conceptual structure of algebra word problems. *Journal of Educational Psychology, 84 (4)*, 419-428.

Wideman, H. H., & Oweston, R. D. (1991, April). *Promoting cognitive development through knowledge base construction.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Yang, C., & Wedman, J. F. (1993). A study of the conditions influencing analogical problem solving. *Journal of Research and Development in Education, 26 (4)*, 213-221.

Table 1 Cell Sample Sizes and Scored Cells

Groups	ICGE 1	ICGE 2	ICGE 3	ICGE 4
1	Concept Maps n = 47	Previous Concept Map N = 47	Previous Concept Map not scored	Outlines not scored
2	Control Not Scored	Concept Maps n = 47	Previous Concept Map n = 45	Outlines not scored
3	Control n=49	Control not scored	Concept Maps n = 39	Outlines not scored
4	Outlines n = 19	Previous Outlines not scored	Previous Outlines not scored	Concept Maps not scored
5	Control Not Scored	Outlines n = 39	Previous Outlines not scored	Concept Maps not scored
6	Control Not Scored	Control n = 44	Outlines not scored	Concept Maps not scored

Table 2 Estimated Task Main Effect Variance Components and Standard Errors

Groups	ICGE 1	ICGE 2	ICGE 3
1	Concept Maps 0.023 (0.025)	Previous Concept Map 0 (0.016)	
2		Concept Maps 0.043 (0.042)	Previous Concept Map 0 (0.008)
3	Control 0.007 (0.008)		Concept Maps 0 (0.019)
4			
5			
6		Control 0 (0.008)	

Table 3 Estimated Subject-by-Task Interaction Effect Variance Components and Standard Errors

Groups	ICGE 1	ICGE 2	ICGE 3
1	Concept Maps 0 (0.065)	Previous Concept Map 0.028 (0.049)	
2		Concept Maps 0 (0.051)	Previous Concept Map 0.141 (0.06)
3	Control 0.177 (0.053)		Concept Maps 0.018 (0.071)
4			
5			
6		Control 0.072 (0.059)	

Table 4: Summary of Task Main Effect Comparisons

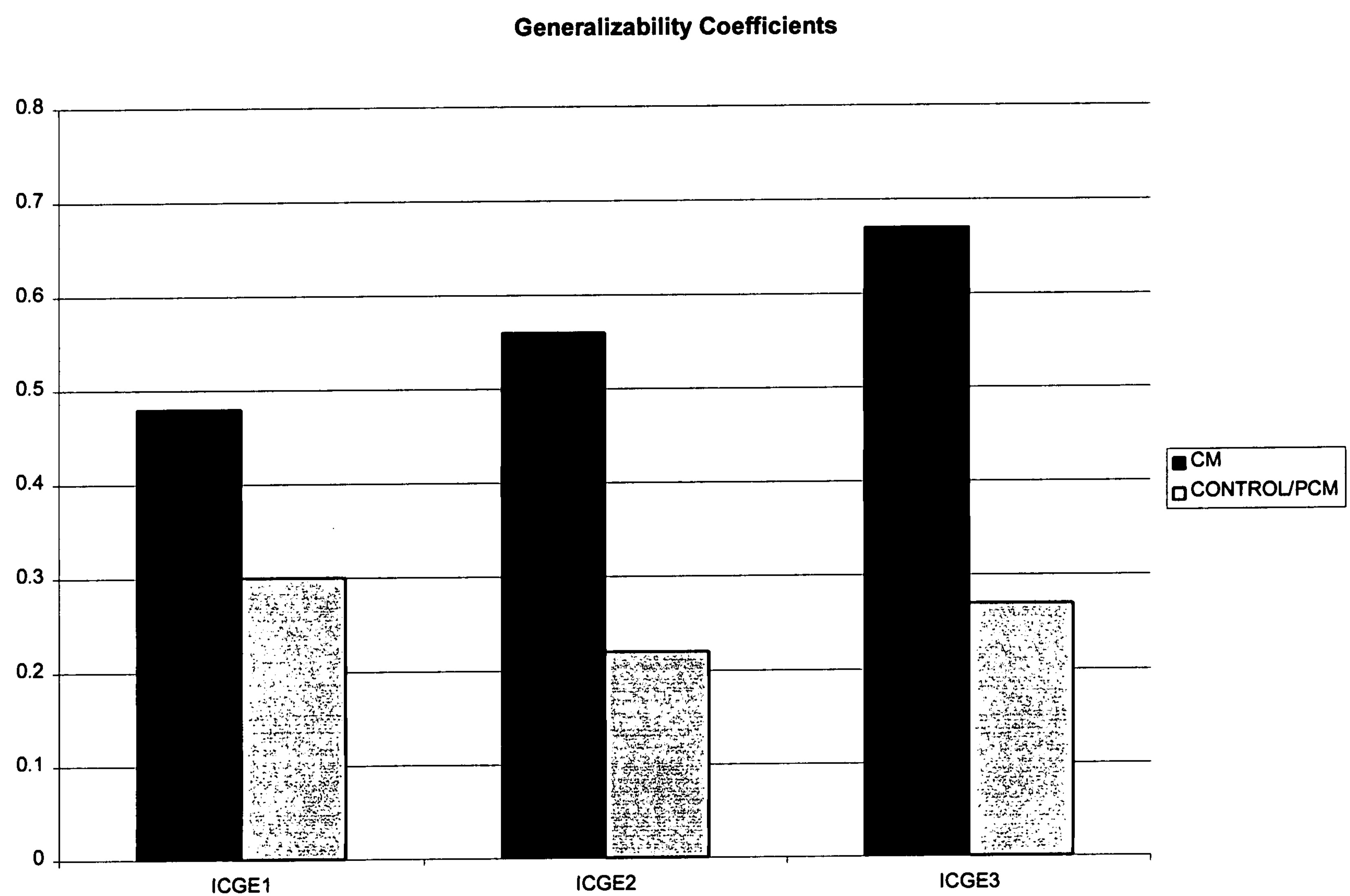
Comparison		Result
ICGE1 Concept Map Condition	ICGE1 Control Condition	n.s.
ICGE2 Concept Map Condition	ICGE2 Control Condition	n.s.
ICGE2 Concept Map Condition	ICGE2 Previous Concept Map Condition	n.s.
ICGE2 Previous Concept Map Condition	ICGE2 Control Condition	n.s.
ICGE3 Concept Map Condition	ICGE3 Previous Concept Map Condition	n.s.
ICGE1 Control Condition	ICGE3 Concept Map Condition	n.s.

Table 5: Summary of Subject-by-Task Interaction Effect Comparisons

Comparison		Result
ICGE1 Concept Map Condition	ICGE1 Control Condition	CM < Control
ICGE2 Concept Map Condition	ICGE2 Control Condition	n.s.*
ICGE2 Concept Map Condition	ICGE2 Previous Concept Map Condition	n.s.*
ICGE2 Previous Concept Map Condition	ICGE2 Control Condition	n.s.*
ICGE3 Concept Map Condition	ICGE3 Previous Concept Map Condition	n.s.*
ICGE1 Control Condition	ICGE3 Concept Map Condition	CM < Control

* These comparisons, though not statistically significant, do trend according to the prediction made in the hypothesis.

Figure 1: Generalizability Coefficients



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029850

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Structural Knowledge as a Pre-requisite to Valid Performance Assessment Scores

Author(s): Jay Parkes, Hoi K. Suen, Dawn M. Zimmaro, Stephen M. Zappe

Corporate Source: University of New Mexico

Publication Date: 4/99

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]

Printed Name/Position/Title: Jay Parkes PhD Asst. Prof.

Organization/Address: 128 Simpson Hall UNM Albuquerque NM 87109

Telephone: 505-877-3820

FAX:

E-Mail Address: parkes@unm.edu

Date: 4-22-99

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com